

How Do We Archive Digital Records?: The Report of the CPA/RLG Task Force

by

Donald J. Waters
Associate University Librarian
Yale University

January 27, 1997

How Do We Archive Digital Records?: The Report of the CPA/RLG Task Force

The Task Force on Archiving of Digital Information

Rapid changes in the means of recording information, in formats for storage, in operating systems, and in application technologies threaten to make the life of information in the digital age much like life in Hobbes' state of nature: "nasty, brutish, and short." The Commission on Preservation and Access and the Research Libraries Group (RLG) created the Task Force on Archiving of Digital Information at the end of 1994 to help relieve building anxiety about the fragility of culturally significant digital information. The Commission and RLG asked the Task Force to frame digital archiving as a set of problems and tasks and to suggest an orderly, perhaps even manageable, approach to their resolution.

The Commission and RLG selected members with a breadth of experience from a broad range of disciplines and backgrounds, including many from the research library community. I am sure that it was an accident, but as if to emphasize the strangeness of the new land they were asking the group to chart, the Commission and RLG selected two co-chairs -- me and John Garrett -- who are both anthropologists by training. In addition to research librarians and anthropologists, the Task Force included archivists, publishers, technologists, bibliographic service vendors, and legal and copyright specialists. The Task Force sponsors then asked the group to seek input from a still wider array of specialists and interested parties by issuing a draft report, distributing it widely, and inviting comment before composing a final report.

The Task Force submitted its draft report in August 1995. The comment period formally ended on October 31, but in fact continued through January 1996. We received numerous thoughtful and helpful comments, suggestions and criticisms from many individuals. The international interest in the report was especially gratifying for those of us on the Task Force. We received extensive comments from a federation of libraries and archival agencies in Australia, from the Library networks and services Directorate of the Commission of the European Union and in January from the Consortium of University Research Libraries in Great Britain.

The Task Force incorporated what we learned into our final report. We corrected the most flagrant errors and infelicities contained in the draft report and, in revisions and an extensive set of annotations, addressed most of the questions and additional issues that arose during the comment period. We completed our work and submitted our final report on May 1, 1996.

The first of nine recommendations that the Task Force made in its final report called for "a cooperative project designed to place information objects from the early digital age into trust for use by future generations." We argued that "action is urgently needed to ensure that documents, software products and other digital information objects that document the early digital age from 1945 to 1990 are preserved before they slip irrevocably away" (Task Force 1996:38). The theme of this meeting -- "Documenting the Digital Age" -- clearly overlaps and advances the interests of the Task Force and I am privileged as a member of that group to be a part of this discussion.

As my contribution to this discussion, I want to stimulate your attention to the question of the means and prospects of digital archiving, to the question of how do we archive digital records? My central argument, following the Task Force report, is that "the problem of preserving digital information for the future is not only -- or even primarily -- a problem of fine tuning a narrow set

of technical variables.... Rather, it is a problem of organizing ourselves over time and as a society to maneuver effectively in a digital landscape. It is a problem of building -- almost from scratch -- the various systematic supports, or deep infrastructure, that will enable us to tame our anxieties and move our cultural records naturally and confidently into the future" (*ibid.*: 6).

To develop our understanding of what is involved in building such a deep infrastructure for digital archiving, I ask you please to join me in thinking through the following chain of reasoning. We need first to accept the premise that archiving is central to the emerging knowledge-based economy. Without economy in the archiving of digital information there can be no real economies in the production and distribution of knowledge. Second, I want to suggest that economies in archiving depend on our understanding of the integrity of digital information objects and arise from the organizational requirements for preserving the integrity of those objects. And finally, I want to persuade you that the path to achieving a knowledge-based economy is actually to set the mechanics of digital archiving in motion as a pervasive and trusted foundation for cultural discourse.

The Value of Archiving in a Knowledge-based Economy

Any discourse about economy, about the efficient management of scarce resources toward valued ends, is ultimately a discourse about values. Of what value or good, we must ask, is archiving and why should we push any scarce resources its way? This is a difficult question about purpose that may immediately open questions about and prompt defenses of particular forms of organization for archiving. In considering the answer, however, we must separate issues of purpose and function from those of organization.

I note in passing here that the Task Force simplified matters greatly in the interest of clarity. It consistently equated long-term preservation with archiving, and its report identifies digital archives, rather than digital libraries, as the unit of activity for the long-term preservation of digital materials. I maintain this usage here and it is a functional, not an organizational distinction. We all know that many libraries do frequently assume responsibility for the long-term preservation of the record of knowledge, but we have come to designate those that exercise such responsibility as a matter of course with special semantic markers as in the phrase "*research library*." Moreover, although we now refer to "digital libraries," discussion of such entities to date has made almost no reference to the long term value of their content nor to the mechanisms that might be employed to preserve such value over time. Rather than use the semantically marked phrase that Peter Graham (1995) has suggested, namely the "digital research library," we adopted the simpler designation of "digital archive."

In answer to the question about the value of archiving, the Task Force report invokes a general "culture-at-risk" argument. Culture -- any culture, so the argument goes -- depends on the quality of its record of knowledge. If that record is defective, as it will be if urgent attention is not given widely to the preservation of information in digital form, then the quality of the culture is also at risk (Task Force 1996: 1-3). The Task Force called attention to the loss of records from the 1960 census, which is a constitutionally-mandated activity, and highlighted other losses of culturally significant records. The Task Force intended its "culture at risk" argument to establish a case for the preservation of digital information as a general matter of public interest and policy.

However, the culture-at-risk argument is a common one, perhaps too common, and is often invoked to attract attention -- and public money or philanthropy -- to a wide variety of issues that otherwise do not carry much economic or political weight. Is there any special force to the argument in this case? There is, I believe, and it emanates from a unique set of factors that are contributing to the emergence and development before our eyes of a broadly-based and powerful knowledge economy. To elucidate these factors, we must identify the principles underlying a

knowledge economy as distinct from other kinds of economy and demonstrate the place of archiving among them. The basic principle that enables us to regard the knowledge economy as a distinct construct is the notion that the pursuit of knowledge is its own end. As I craft for your review an analysis of the special force of the culture-at-risk argument based on this fundamental principle, I turn for help to two unlikely sources: the works of Richard Lanham, a distinguished professor of English at UCLA, and of Jaroslav Pelikan, the great religious historian at Yale.

In *The Electronic Word*, among other recent publications, Richard Lanham (1993) has argued that the scarce commodity in a knowledge-based economy is not information. We are glutted with information. Rather, the scarce commodity is the human attention which gives information its structure, its usefulness and its value as knowledge. In Lanham's scheme, human attention is labor, information technology is the means by which the labor is applied, and attention-structures designed to capture the interest of consumers, including students and other scholars, are the products of the labor and the technology.

At its core, Lanham's theory is an application of the labor theory of value to the knowledge economy. His unique contribution to the theory is his further argument that the discipline of rhetoric provides the theoretical framework for systematically describing and evaluating the end products of this knowledge work, the attention-structures. Note, however, the distinctive quality that Lanham attributes to the knowledge economy: as attention-structures, or works of knowledge, capture attention, they beget further attention-structures. Knowledge begets knowledge; knowledge is its own end.

In *The Idea of the University: A Reexamination*, Pelikan (1992) has produced one of the most eloquent and detailed critiques of the principle of knowledge as its own end and of the university, in which the principle has long provided the central operating concept. According to Pelikan, the principle of knowledge as its own end is merely one of a more comprehensive set of first principles that he calls the "intellectual virtues." These virtues are essential for the development of knowledge, and include principles of free inquiry and intellectual honesty, an obligation to convey the results of research, and an affirmation of the continuity of the intellectual life, upon which each generation builds and to which it contributes in turn (*ibid.*: 32-56). Building on this set of first principles, Pelikan argues that the advancement of knowledge through research, the transmission of knowledge through teaching, the diffusion of knowledge through publishing, and the preservation of knowledge in scholarly collections are the four legs supporting any table made for the pursuit of knowledge; they particularly support the table that has come to be known as the research university (*ibid.*: 16-17, 78-133).

Invoking the 19th century phrasing of John Henry Newman, Pelikan goes on to suggest that support for teaching, research and publication constitutes the "endowment of living [genius]" while efforts to preserve, or archive, knowledge by organizations like libraries, museums and archives, represent "the embalming of dead genius" (*ibid.*: 110). Lest the connotations of these archaic phrases give you pause, note that Pelikan is careful to distinguish embalming from entombing and his use of "embalming" is a colorful synonym for preservation and archiving which he takes to include all of the means necessary to make knowledge accessible to present and future generations. Moreover, he vigorously argues that "new knowledge has repeatedly come through confronting the old, in the process of which both old and new have been transformed" (*ibid.*: 120). Memory is not a warehouse, but an active process of re-categorizing based on previous categorizations. In the province of the knowledge economy that we know as the research university, the two motives at work -- embalming and endowment of genius, the looking backward in preservation and the looking forward in research, teaching and publication -- thus are inextricably linked and flow from the principle that the pursuit of knowledge is its own end: preserved work from past generations is a necessary foundation for present and future work, which in turn defines the accessibility of the preserved work.

If we accept the argument that the emerging knowledge economy is founded on the principle of knowledge as its own end and that the broadly defined function of preserving, or archiving, the record of knowledge is essential to the pursuit of knowledge, then it follows that the emerging knowledge economy cannot survive without a provision for the archiving function. It is this logic that led the Task Force to assert as a fundamental principle that "information creators/providers/owners have initial responsibility for archiving their digital information objects" (Task Force 1996: 20). In a knowledge economy, where knowledge is both the source and outcome of labor, we presume archiving to be in the producers' own self-interest.

Did we on the Task Force believe that such self-interest is sufficient in all domains to meet the requirements of a larger knowledge-based culture? Not at all. The knowledge economy is in very early stages of development and such a virtuous outcome is by no means secure.

The developing nature of the knowledge economy is palpable. We all feel it and we do, in part, because one of its fundamental characteristics is the force of democratizing the value of knowledge as its own end. Lanham marshals considerable evidence that the rapid expansion of the division of labor around digital technologies -- what George Gilder (1995; see also Bronson [1996]) calls the technologies of sand (for silicon chips), glass (for optical networks) and air (for wireless networks) -- has democratized the cultures it touches. The products of the knowledge economy -- the attention-structures -- are easier to generate and to use. They make knowledge more accessible. The markets for them continue to expand demanding more knowledge workers and creating more knowledge consumers who are broadly educated in the arts and sciences.

In the US and elsewhere abroad, the pressure of these developments on the educational system, particularly the system of higher education, has been extraordinary and, at least, fourfold. First, the system must serve a growing number of students who by conventional standards need remedial training to advance through the curriculum. This pressure is, in part, an expression of the distinction between the *haves* and the *have-nots*. Second, the broader range of constituents in higher education, whether they need remedial training or not, presses for different approaches to the curriculum. The expression of this pressure appears, in part, in the form of debates over the place of multiculturalism on our campuses and in our curricula. Third, the division of labor in the knowledge economy has resulted in both increasing specialization within disciplines *and* the rapid growth of interdisciplinary study. And fourth, the system can only serve the broader range of constituents and interests by dramatically lowering the costs of education to affordable levels.

Note, as Lanham does, that the dynamic described here represents a profound impulse to achieve the preeminent goal of education in a democracy (Lanham 1993: 23). That is, insofar as knowledge is both the source and outcome of human labor in this rapidly growing segment of the economy, literate citizens will prevail who value the lifelong pursuit of knowledge as its own end, as both the source and outcome of their labor. Yet, ironically, just as this democratic goal has come into plain view, what Donald Norman (1988: 1-33) calls the "psychopathology of everyday things" intrudes. Despite the declining costs of information technologies, long-promised productivity gains remain elusive, especially in higher education. Absent such gains, the cumulative result of the social and economic pressures to lower the cost of education feels to many of us in the business as if we are under siege and being asked simply to lower the quality of our products and services.

Following Norman, I submit that the solution to the productivity paradox in the knowledge economy is a matter of design and development. Because digital information is not only a product but is also a source of knowledge, it will remain difficult and costly to use as long as its design makes it difficult or costly to maintain use, especially over the long term. Economy in the use of digital information, in other words, requires an economy of digital archiving.

Developing an economy for digital archiving

If preservation is an essential feature of the knowledge economy, then real economies are necessary and must emerge in digital archiving for the knowledge economy truly to flourish. Posed in this way as a problem of economic development, those of us in the business of managing the record of knowledge for posterity can easily succumb to terror in the face of the explosion of digital information. I identify my own feelings with those of the woman so clearly captured in one of James Thurber's wonderful cartoons. She sits before a bow-tied, pince-nezed, and long-eared doctor to whom she has come for help. He looks just like a rabbit and he observes: "You said a moment ago that everyone you look at seems to be a rabbit. Now just what do you mean by that, Mrs. Sprague?" (Grauer 1995: 148). Everywhere we look, there is digital information. How do we put ourselves as a culture in the position of identifying and giving sufficient attention to the digital material that is worth saving?

Although the task overall may be daunting, we are not helpless and without places to start. Observe, for example, that the real intellectual action for at least a subset of scholarly disciplines no longer even occurs in the conventional publication stream but elsewhere: in on-line databases, on-line exchanges of pre-prints, listservs and so on. Conventional publication in these disciplines adds little value to the work that has already been disseminated in these other channels; rather it is a redundant process, undertaken to generate, in effect, a certified archival record of the work. Because the audience paying attention to the field has already seen and absorbed the work in on-line versions, the printed publication channel grows increasingly narrow consisting primarily of libraries who serve as the archival institutions. Given a narrow market, costs and prices consequently rise on the supply side. On the demand side, libraries respond by cutting titles from their collections (Waters 1996).

There is clearly little logic or economy in a process whereby scholars use printed publications to establish an archival record only to find that the institutions responsible for ensuring that the archive endures for future generations cannot afford to purchase the publications. Framed in this way, the problems in the scholarly communication process that appear to us as a spiral of escalating prices and journal cancellations are *archival* problems. As such, they give research libraries, publishers, scholars and universities substantial economic motive to save money and streamline the process. Where there is redundancy between print and electronic form, as there increasingly is in disciplines such as mathematics and physics where pre-print markets flourish, we need to identify and capture the real intellectual activity from the on-line places wherever it is now naturally occurring and ensure that such activity is housed in certified, durable and readily accessible archives. In so doing, we can eliminate a substantial set of redundant costs and perhaps even enable our colleagues in the academy to change further the ways in which they conduct scholarship and also, perhaps, the mechanisms, such as tenure review, by which they measure the quality of that work.

We will do ourselves and our colleagues no favors, however, if we replace a costly, redundant archival process with one for digital materials that is even more costly. The Task Force (1996: 9-36) identified a wide range of factors the interaction of which provides fertile ground for the development of economies in archiving. The factors include the various kinds of digital information objects -- text, images, numeric data, sound, video, simulations, geographic information systems, hypermedia and so on -- and the various claims of stakeholders with interests in the creation, management, dissemination, use and retention of digital information. Perhaps the most significant factors are those affecting the integrity of information objects in whatever form they may appear, and those required specifically for the organization of archives.

The integrity of information objects. The central goal of preservation must be to preserve the integrity of the object. Knowing how to preserve a digital information object depends on being

able to define and preserve the features that give it a distinct identity and define it as a whole and singular work. In the digital environment, the features that determine information integrity and which deserve special attention for archival purposes include the following: content, fixity, reference, provenance, and context. Choices about each of these features significantly affect the economy of archiving.

Choices about preserving the content of digital information objects range over a continuum of abstraction. At the lowest level of abstraction, preserving content simply means preserving a collection of bits. An archival choice at this level often means preserving the hardware and software that may be uniquely capable of interpreting the bits associated with a particular information object. Preserving content may also refer to preserving the composition of ideas in a particular structure and form. Encoding characters in ASCII or UNICODE provides varying ability to represent multiple languages, and formula and equations. Markup languages, such as TeX, SGML and HTML, offer both advantages and disadvantages in representing layout and document structure compared to the use of proprietary word processing systems and interchange formats. In the realm of digital images, consideration of resolution, color and compression often pits the quality of content representation against storage efficiency and loss of content. Finally, preserving content may refer, at the highest level of abstraction, to preserving the knowledge and ideas embodied in an object in a way that transcends the limits of the hardware and software needed to read bits or to render the information for use in a specific format or structural representation.

Preserving the fixity of information objects is especially troublesome in the digital world, where objects are frequently subject to change or withdrawal. Outside the digital arena, there are various methods of fixing information in objects: business records contain evidence of transactions, the acts of production and broadcast record specific radio and television programs, and publishers generate specific versions or editions of works. In the digital environment, however, the use of cryptography and other techniques is still maturing to support digital archives in establishing trusted channels of distribution, and to help them discriminate among multiple versions and to identify canonical versions. Moreover, some digital information objects are better modeled as continuously updated databases for which the preservation choice is whether to compile a complete record of changes or to capture snapshots of the database as the means of preserving information integrity.

Systems of citation, description, and classification provide the necessary means of reference for consistent discovery, identification, and retrieval of information objects over time. Preserving reference is thus an essential means of preserving the integrity of digital information, but it is problematic for several reasons. Self-referential information in digital objects seldom meets conventional citation quality. Moreover, consistently resolving names and locations of digital objects is, given the current state of the art, either difficult or unreliable. Finally, conventional reference mechanisms, such as on-line catalogs, do not easily accommodate certain kinds of reference data, such as information about the terms and conditions of licenses for intellectual property, which increasingly govern the use and cost of culturally-significant information objects in the digital world.

Provenance is another essential feature of information integrity, and refers to the origin and chain of custody through individuals, organizations and instrumentation, including within the archive itself. By documenting provenance, archives create the presumption that an information object is authentic. Compared to conventionally published objects, which employ well-known techniques for establishing their origin that are usually shown on a title page or its verso, the means of establishing the provenance of information published digitally are not yet well established. In addition, there are special problems in the digital world, as in other arenas, for establishing the provenance and authenticity of individual records, such as mail, diaries and personal databases,

and of corporate records, the understanding of which depends fundamentally on an appreciation of their origins in policies, procedures, and organizational roles and responsibilities. Of special note are the integrity problems associated with digital information objects produced by digital instrumentation in scientific experiments, clinical services and remote sensing. Establishing provenance of these objects -- and thus their integrity -- requires a detailed understanding of the calibration, units of measure, sampling rate, recording conditions, and other features of the instrumentation that generated the information (see National Research Council: 1995a,b).

The fifth attribute of information integrity that bears on the preservation of digital information objects is their context, the ways in which they interact with elements in the wider digital world. Among the various dimensions of interaction, there is a technical dimension, in which digital objects depend for their existence on specific hardware and software. There is also a dimension of linkages to other objects. In the World Wide Web, the integrity of many objects resides in the network of linkages. To preserve both the objects and the linkages is a daunting challenge for which there exists no good solution today other than to take periodic snapshots of the network objects. A communications dimension of information context defines the effects of the medium of transmission, such as CD-ROM or networks of varying bandwidth, on the types and characteristics of digital information objects. Finally, a social dimension, in which government policies, role relationships, and other political and organizational factors shape the creation and use of digital objects, also affects information integrity and the ability of archives economically to preserve it.

The organization of archives. Another set of factors that the Task Force on Archiving of Digital Information identified as grounds upon which to develop an economy of digital archiving is the set of factors required specifically for the organization of archives. The digital environment today is so fragile that those who disseminate, use, re-use, re-create, and re-disseminate various kinds of digital information can easily, even inadvertently, destroy valuable information, corrupt the cultural record, and ultimately thwart the common pursuit of knowledge. Digital archives build and maintain reliable collections of well-defined digital information objects and they preserve the features -- content, fixity, reference, provenance, and context -- that give those objects their integrity and enduring value. They do so by managing costs and finances within an operating environment that has a core set of features including the means of migrating digital information to maintain its vitality as hardware and software environments change.

Among the core set of features in the operating environment of digital archives is a selection and appraisal process. Archives cannot save everything. To identify the most valuable objects for preservation, archives must appraise the content of the object -- its subject and discipline -- in relation to the collection goals of the digital archives, the quality and uniqueness of the object, its accessibility in terms of available hardware, software and legal status, its present value, and its likely future value. Once an object is selected for inclusion, it needs to be accessioned -- that is, prepared for the archives. Accessioning involves describing and cataloging selected objects, including their provenance to authenticate them, and securing them for storage and access. Storage, depending on expected use and the kind of performance needed in retrieval, may be on-line in magnetic media, near-line in optical or tape media in a jukebox retrieval system, or off-line in media that requires manual intervention to retrieve. Access systems must facilitate discovery, retrieval, and use, including the management of intellectual property rights as appropriate, in a distributed, presumably networked, environment. Finally, digital archives need a high level of systems engineering skill to manage the interlocking requirements of media, data formats, and hardware and software, and to help determine when objects should migrate to new systems or system components.

Migration is the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent

generation. As the Task Force defined it, "the purpose of migration is to preserve the integrity of digital objects and to retain the ability for clients to retrieve, display and otherwise use them in the face of constantly changing technology" (1996: 5). Digital archives have various migration strategies available to them. Internally, they can build hardware or software emulators to preserve the technical operating environments of the information objects, they can change, or "refresh," the media on which the objects are stored as storage technology evolves, or they can reformat the objects to accommodate changing technology. In addition, they can work externally with creators so that digital information incorporates standards that simplify the migration issues. They can work with systems designers to engineer cost-effective migration paths into the hardware and software on which information objects depend. Finally, they can use processing centers that develop best practices and achieve economies of scale in certain kinds of migration techniques.

Means of economizing. In this complex mix of factors by which digital archives operate to preserve the integrity of digital information objects, there is much room for the play of specialization, division of labor and competition that will not raise costs, but drive the economy of archiving vigorously to lower them. Division of labor and specialization are already evident. For example, some key services, such as rights management and network charging facilities, are emerging generally in the commercial marketplace and will undoubtedly serve the interests of archiving as well as other segments of the knowledge economy. The development of other services, such as durable naming conventions and expanded metadata facilities, are well underway. Still other kinds of specialized archival services -- those, for example, that require the complex weaving of information holdings in particular disciplines from among a variety of providers and custodians -- will require time and a commitment to a complex iteration and reiteration of exploration, development and solution as the relevant issues emerge and become clearer and more tractable.

Fortunately, as we design these explorations, we have a rich experience from which it behooves us to draw. In the creation of information utilities like OCLC and the Research Libraries Group, libraries in the US came together to craft an economy out of an information management process in which we had formerly operated handicraft style in isolation and without the discipline of competition and the benefits of economies of scale. Just as we did two decades ago for bibliographic control, we have to find ways to invest our interactions over digital archiving with a marketplace dynamic that drives us to organize and routinize the activity and thereby continually to improve quality and lower cost.

The process of coming to terms with each other, and with our partners in academia, in publishing and in the larger knowledge economy about the investments we must make in digital archiving is essentially a coming to terms about the centrality of archiving -- the embalming of dead genius -- in the pursuit of knowledge. But these understandings and agreements can be achieved only in actual practice. And this brings me to my third and final point: that our agreements to divide the labor as formal partners, as informal allies, even as competitors must set in motion soon and substantially the mechanics of digital archives as a pervasive and trusted foundation for cultural discourse.

The Mechanics of Digital Archives

There is an apocryphal story about the government service agency that formulated its record retention rules as follows: 1) discard all records when they become 30 years old; 2) retain all records over 50 years old, for their historical value (National Research Council, 1995b: ix). Most of the Task Force recommendations are designed explicitly to avoid the paralysis of this kind of thinking about the emerging knowledge economy. The recommendations for setting in motion the mechanics of digital archiving invite substantial cooperative action. They are grouped in three categories: pilot projects which focus on content, technologies, and the legal and economic

barriers to archiving; support structures, including national policy, legal and institutional foundations for fail-safe mechanisms, notions of certification, scholarly and professional societies who need to worry about archiving the digital objects they care most about, and an international point of contact; and best practices for supporting archiving at the point of creation, for storage, for discovery and retrieval, and for migration. I draw your attention to three of the recommendations. They each illustrate a different form of interaction and they each yield a different kind of benefit.

First, the Task Force called for certified digital archives. The process of certification is meant to create an overall climate of value and of trust about the prospects of preserving digital information. Repositories claiming to be digital archives in a changing and uncertain environment must be able to prove that they are who they say they are, and that they can deliver on the preservation promise. There are at least two models of certification. On the one hand, there is the audit model used in the US, for example, to certify official depositories of government documents. The depositories are subject to periodic and rigorous inspection to ensure that they are fulfilling their mission. On the other hand, there is a standards model which operates, for example, in the preservation community. Participants claim to adhere to a given set of standards; consumers certify by their use whether the products actually adhere to the standards. The Task Force did not judge the merits of these alternatives. Instead, its call for individuals and organizations to agree to collaborate in the design and implementation of standards, criteria, and mechanisms for certification, and for prospective digital archives to submit to the certification process is a summons for the wider community to affirm the values -- at least in the abstract -- of digital preservation and ultimately of the pursuit of knowledge as its own end.

The Task Force also emphasized the need for a fail-safe mechanism in digital archives. Such a mechanism will enable a certified archival repository to exercise an aggressive rescue function to save digital information that it judges to be culturally significant and which is endangered in its current repository. We may not know enough about the use of digital information to reach consensus just yet about what fair use of it is, but we do know that one of the greatest dangers to its long life is the ease with which it can be abandoned or destroyed. If concerted action is needed in the intellectual property arena to balance the rights of creators and publishers against the need to support teaching and research, then let us focus at least some of that action on the development of the legal framework needed to support a fail-safe mechanism for digital archives. The benefit of such action is, of course, not in the dollars it directly generates or saves, but in the environment it creates for archival institutions to do their job and to realize the value of preserved work for future generations.

Finally, I call attention again to the overlapping interest of those at this meeting with the Task Force, whose members recommended a cooperative venture to preserve the documents, discourse, software products and other digital information objects that serve to record the early digital age. Because the objects in this focal area are at such risk of loss, the project could provide a useful means of exploring the actual operation of archival fail-safe mechanisms. Moreover, conceived as a cooperative venture among multiple participating archives, the project would provide a necessary testbed for developing an on-line system of linked but distributed archives. One of the biggest unknowns in the digital environment is the full impact of distributed computing over electronic networks. However, as the Task Force report suggests in the section on costs and finances, one of the greatest hopes for reducing costs in the scholarly communication process is the prospect of achieving economies of scale in the storage and distribution of electronic information over electronic networks. We need to verify these expectations of economic benefit in actual experience with a range of materials in archival settings.

Conclusion

I conclude by observing that the notions of archives and archiving today have much currency and import, even outside the context in which we have been discussing them here. Such currency is evidence, perhaps, of the democratizing effects of the knowledge economy. In the *New York Times Magazine* at the end of 1995, William Safire devoted one of his "On Language" columns to the topic of kids' slang. He advised that "if you want to stay on the generational offensive, when your offspring use the clichéd *gimme a break*, you can top that expression of sympathetic disbelief with *jump back* and the ever-popular riposte *whatever*." However, he noted that some expressions, such as *I'm outta here* or *I'm history*, are now very much dated. *I'm history*, Safire quoted a forthcoming study of slang, is "a parting phrase modeled on an underworld expression referring to death" -- remember what I said about embalming -- and the phrase has both inspired and been replaced by the more trendy expression, *I'm archives* (Safire 1995: 30).

With regard to the future of digital information in the pursuit of knowledge, I have no doubt that the expression *I'm archives* will apply truthfully to all the institutions represented in this conference. The choice before us, both individually and collectively, is to decide in what sense it will apply.

References

Bronson, Po

1996 "George Gilder." *Wired* 4.03 (March): 122-126, 186-195.

Gilder, George

1995 "Angst and Awe on the Internet." *Forbes ASAP* (December 4).
Available: [WWW:<http://homepage.seas.upenn.edu/~gaj1/ggindex.html>].

Graham, Peter

1995 "Requirements for the Digital Research Library." *College and Research Libraries* 56(4): 331-339.

Grauer, Neil

1995 *Remember Laughter: A Life of James Thurber*. Lincoln: University of Nebraska Press.

Lanham, Richard A.

1993 *The Electronic Word: Democracy, Technology and the Arts*. Chicago: The University of Chicago Press.

National Research Council

1995a *Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources*. Washington, D.C.: National Academy Press.

1995b *Study on the Long-Term Retention of Selected Scientific and Technical Records of the Federal Government: Working Papers*. Washington, D.C.: National Academy Press.

Norman, Donald

1988 *The Psychology of Everyday Things*. New York: Basic Books, Inc.

Pelikan, Jarislov

1992 *The Idea of the University: A Reexamination*. New Haven: Yale University Press.

Task Force on Archiving of Digital Information

1996 *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information*. Washington, D.C.: Commission on Preservation and Access and Mountain View, CA: The Research Libraries Group, May 1.
Also available: [WWW:<http://www-rlg.stanford.edu/ArchTF>].

Safire, William

1995 "Kiduage." *The New York Times Magazine*. October 8, 1995: 28, 30.

Waters, Donald J.

1996 *Realizing Benefits from Inter-Institutional Agreements: The Implications of the Draft Report of the Task Force on Archiving of Digital Information*. Washington, D.C.: The Commission on Preservation and Access.
Also available: [WWW:<http://arl.cni.org/arl/proceedings/127/waters.html>].